



# Digital Government Academy Course: Enterprise Content Management

Presented by Interwoven, Inc, for the State of Washington

# Today's Agenda

- ◆ Today we'll continue our exploration of Enterprise Content Management in the State of Washington

- ◆ Today's topics:

- **TeamSite and Metadata**
- Finding your Assets
- Supporting Personalization



- ◆ **Objectives:**

- Metadata concepts
- Metadata technologies
- Metadata planning and analysis

- ◆ **Action planning:**

- Analyzing the metadata requirements for the DSHS sample project
- Analyzing metadata requirements for your own agency

# Today's Agenda

- ◆ Today we'll continue our exploration of Enterprise Content Management in the State of Washington

- ◆ Today's topics:

- TeamSite and Metadata
- **Finding your Assets**
- Supporting Personalization



- ◆ **Objectives:**

- Development environment searching
- Production Environment searching

- ◆ **Action planning:**

- Analyzing your development and production environment search requirements, with possible changes to your metadata planning

# Today's Agenda

- ◆ Today we'll continue our exploration of Enterprise Content Management in the State of Washington

- ◆ Today's topics:

- TeamSite and Metadata
- Finding your Assets
- **Supporting Personalization**



- ◆ **Objectives:**

- Personalization planning

- ◆ **Action planning:**

- Analyzing your personalization requirements

# Seminar Schedule

- ◆ **Day 1: December 4, 2002**
  - Course Kickoff
  - ECM
  - TeamSite Templating
- ◆ **Day 2: December 11, 2002**
  - Develop-and-Deploy
  - TeamSite in the Enterprise
  - Branching Structures
- ◆ **Day 3: December 18, 2002**
  - TeamSite Security
  - Workflow Scenarios
  - Designing a Workflow
- ◆ **Day 4: January 8, 2003**
  - Designing Data Capture Forms
  - Designing Presentation Templates
  - Re-use via Templating
- ◆ **Day 5: January 15, 2003**
  - TeamSite and Metadata
  - Finding your Assets
  - Supporting Personalization
  - Course Summary



## TeamSite and Metadata

# Topic Objectives

- ◆ In this topic, we'll cover:
  - Metadata concepts
  - Metadata technologies
  - Metadata planning and analysis
- ◆ The action planning at the end of this topic will be:
  - Analyzing the metadata requirements for the DSHS sample project
  - Analyzing metadata requirements for your own agency

# What is Metadata?

- ◆ Metadata is "data about data"
- ◆ This includes all information used to maintain content files, including:
  - Categorize
  - Track
  - Search
  - Organize
- ◆ Metadata is valuable both during development and production
- ◆ Different types of content assets will require different metadata
- ◆ Each version of a file can have its own unique set of metadata



## Metadata: Example

- ◆ For website press release HTML files, the following metadata might be relevant:
  - Date of publication
  - Original author
  - Approving reviewer names and dates
  - Cross-reference to other press releases
  - Date of expiration
  - Abstract and keywords
  - Distribution keys (i.e., what audience is the release for)

# Metadata Storage

- ◆ Metadata is stored in the TeamSite backing store separately for each file version
- ◆ The metadata is not inserted into the file, nor is the file's contents changed by adding or changing its metadata
  - However, adding or changing metadata on a file does create a new version

# Metadata Analysis

- ◆ To plan for metadata implementation, consider your organization's content tracking and usage needs
  - Do you plan to automatically expire content?
  - Do you need to track the creation of content from its source through to final publication?
  - Who are the interested parties in your organization who need the data, and what do they do with it?
  - Do you have a problem with search?
  - Do you have an application other than search engines that might consume metadata?

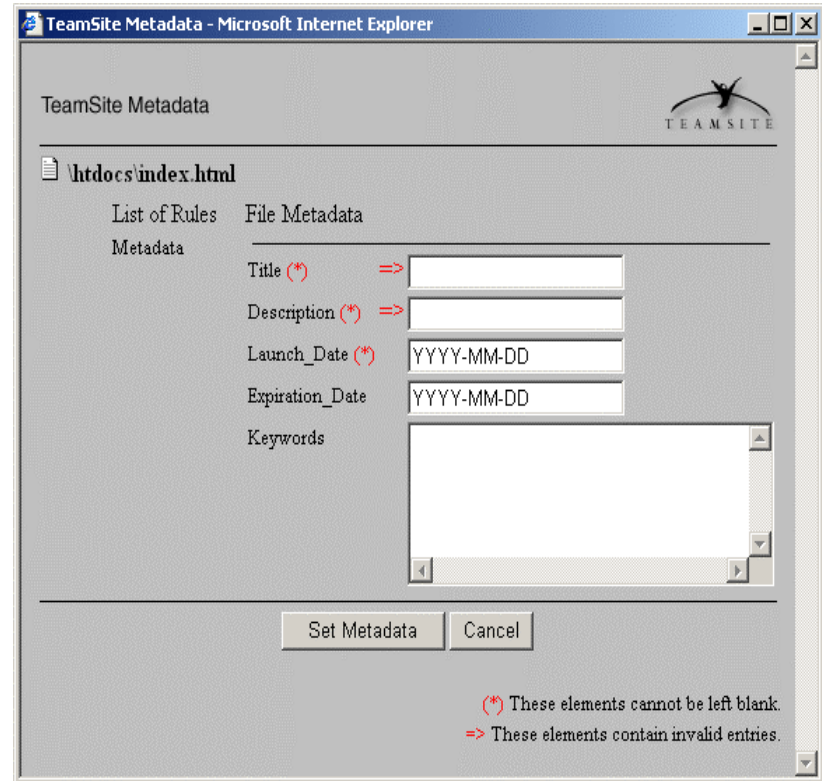
# Interwoven Metadata Technologies

- ◆ There are two technologies available to Interwoven customers for metadata storage and tracking
  - **TeamSite Metadata Capture**
    - Provides basic metadata capture and storage capability
    - Built into TeamSite—always available
  - **Interwoven MetaTagger**
    - Provides advanced metadata capabilities
    - Separate license

# TeamSite Metadata Capture

## ◆ Capabilities:

- Each file can have an arbitrary list of metadata items attached, as name-value pairs, or **tuples**
- Each branch, sub-directory, and/or file type can have its own unique set of metadata tuples
- The metadata capture form is designed in the same way as a TeamSite Templating data capture form
- The submit workflow can be customized to require the metadata capture form to be completed as part of the submit process



TeamSite Metadata

htdocs\index.html

List of Rules    File Metadata

Metadata

Title (\*) =>

Description (\*) =>

Launch\_Date (\*) YYYY-MM-DD

Expiration\_Date YYYY-MM-DD

Keywords

Set Metadata    Cancel

(\*) These elements cannot be left blank.  
=> These elements contain invalid entries.

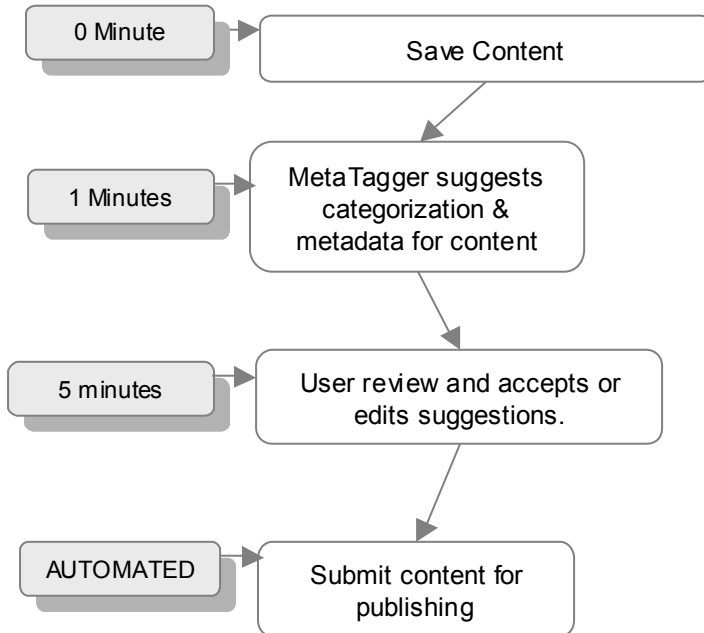
# MetaTagger

## ◆ Capabilities:

- All the capabilities of TeamSite metadata capture
- Automated tagging based on automatic analysis of content
- Intended to support search and classification systems
- Supports both individual file and batch metadata tagging
- Out of the box MetaTagger provides keyword extraction and summarization
- By using a vocabulary (either a custom or one of the provided vocabularies), an enterprise can map out its intellectual property

# Single Asset Metadata Tagging – Process Improvement

## With IWOV MetaTagger



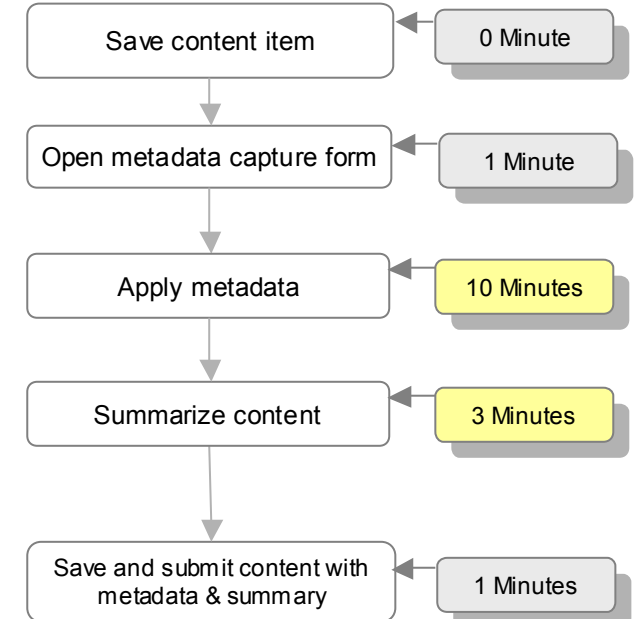
Cost of Labor = \$1 per minute  
Total Time = 6:00 minutes

**Cost per page = \$6**

Savings Per Page = \$9  
Pages = 100,000

**Total Savings = \$900,000**

## Without IWOV MetaTagger

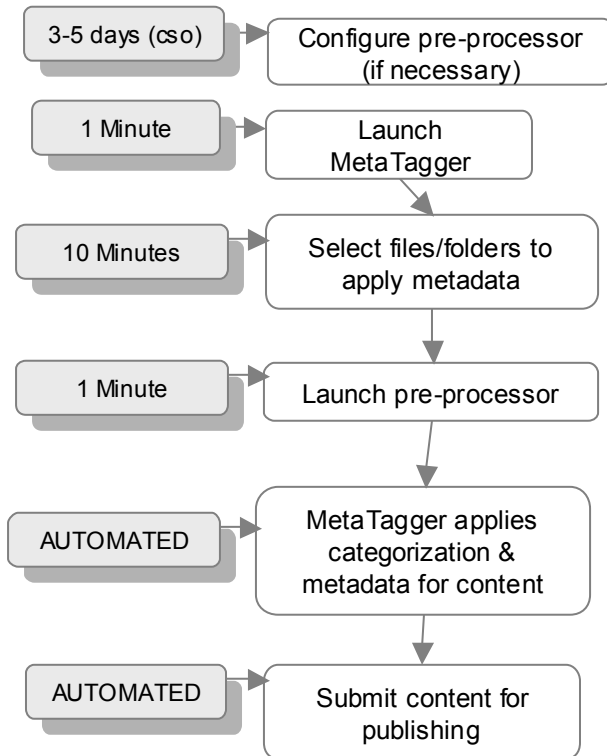


Cost of Labor = \$1 per minute  
Total Time = 15:00 minutes

**Cost per page = \$15**

# Batch Process Metadata Tagging – Process Improvement

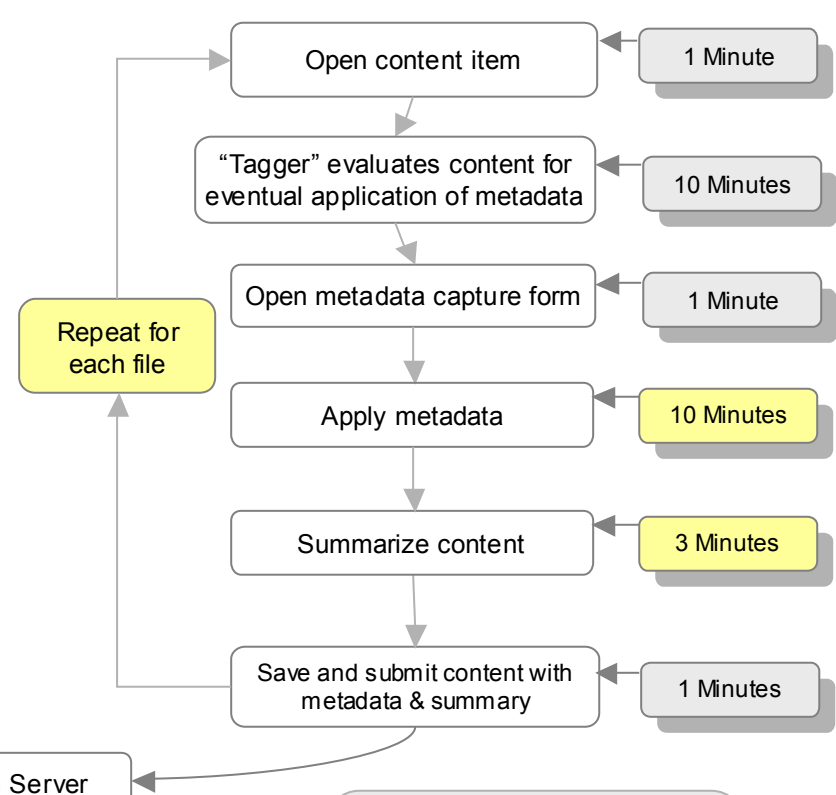
## With IWOV MetaTagger



Cost of Labor = \$1 per minute  
Total Time = 11:00 minutes  
Consulting = \$9k-\$15k

**Cost per page = \$11.00**

## Without IWOV MetaTagger



Cost of Labor = \$1 per minute  
Total Time = 15:00 minutes

**Cost per page = \$15**

**Savings Per Page = \$9**  
**Pages = 100,000**  
**Consulting Services = \$15,000**  
**Total Savings = \$885,000**



# Deploying Metadata

- ◆ To be used, metadata must be deployed to a production database server
- ◆ For instance, metadata for use in search systems must be deployed to the online database so that it can be queried by end users to locate files
- ◆ Each metadata record deployed must contain a reference to the source file, as well as columns for each metadata tuple
- ◆ Interwoven DataDeploy automatically deploys metadata into your own database server's format
  - DataDeploy can also deploy templating data records in the same way
- ◆ Typically, metadata is deployed to the database server when new files are submitted

## Action Exercise: Metadata Analysis

- ◆ Within your group, spend 30 minutes discussing the following:
  - Possible metadata needs for the DSHS press release sample project
  - Possible metadata needs for your own agency
- ◆ After discussion, spend 20 minutes documenting the following:
  - Your list of metadata tuples for the DSHS project
  - How the metadata will be used in development and in production
  - If you have time, also document the same information for your own agency
- ◆ Each group will then present their findings to the class one at a time

# Action Item Discussion

- ◆ Class presentation
- ◆ Questions
- ◆ Take a few minutes to consider other group action items
  - Integrate theirs with yours if needed



## Finding Your Assets

# Topic Objectives

- ◆ In this topic, we'll cover:
  - Development environment searching
  - Production Environment searching
- ◆ The action planning at the end of this topic will be:
  - Analyzing your development and production environment search requirements, with possible changes to your metadata planning

# Searching in the Development Environment

- ◆ In a large TeamSite backing store it can be difficult to locate a particular file quickly
- ◆ Development times can be shortened by enabling a search tool for the development environment
- ◆ If your metadata is being deployed to a database server, you can enable **Metadata Search** in TeamSite
- ◆ Metadata search is a powerful form-based query tool that enables users to search for and locate files in a large TeamSite backing store based on their metadata values
- ◆ You can also search for template data content records based on their item values

# Setting Up Metadata and Templating Search

- ◆ To use metadata search, you must:
  - Configure metadata capture for files in your backing store
  - Install and configure TeamSite Templating
  - Install Interwoven DataDeploy
  - Configure DataDeploy to deploy metadata and/or template DCR records to a database server
  - Install the `iwsearchmetadata.cgi` and `iwsearchdcr.cgi` search tools
- ◆ Instructions in DataDeploy and TeamSite Templating product manuals

## Searching in the Production Environment

- ◆ Once the content has been deployed from the TeamSite server to a production web server, you can also deploy the metadata from TeamSite to a production database server to serve as a search base for real-time queries by users
- ◆ TeamSite and MetaTagger can both generate the data for a search database, and it can be deployed to your database server with DataDeploy
- ◆ You must provide the actual search software and user interface



## Example: Integration with Inktomi

- ◆ When OpenDeploy is run a process extracts the metadata from the files being deployed
- ◆ This metadata is then deployed to the Inktomi search engine
- ◆ The files are then deployed to the production web servers
- ◆ The integration calls Inktomi's API to add the metadata to the search engines collection and tell Inktomi to examine the list of files deployed
- ◆ Now the files are immediately searchable

## Action Exercise: Search Planning

- ◆ Within your group, spend 20 minutes discussing the following:
  - What production metadata from the previous exercise would need to be searchable?
  - What types of search operations do you need to support on your production web site?
  - After discussion, spend 20 minutes documenting the following:
    - Your 5 most common development search criteria
    - Your 5 most common production search criteria
    - If needed, amend your metadata plan with new fields based on this analysis
- ◆ Each group will then present their findings to the class one at a time

# Action Item Discussion

- ◆ Class presentation
- ◆ Questions
- ◆ Take a few minutes to consider other group action items
  - Integrate theirs with yours if needed



## Supporting Personalization

# Topic Objectives

- ◆ In this topic, we'll cover:
  - Personalization planning
- ◆ The action planning at the end of this topic will be:
  - Analyzing your personalization requirements

# Personalization and TeamSite

- ◆ Personalization support is provided via metadata
- ◆ Your metadata plan must include data used for personal preferences and customizations
- ◆ TeamSite does not include personalization runtime software
  - It is used to capture and store the personalization data
  - DataDeploy is used to push the personalization data to the production server
  - At that point, proprietary runtime personalization server software must provide the actual personalization routines and interfaces

# Possible Personalization Factors

- ◆ Typical personalization services might provide:
  - Customization of portal home page
  - Language
  - Format (font, size, color, etc)
  - Filtering of subjects
  - Stored search patterns
  - Personalization – Showing me thing based on settings I chose
  - Customization – Showing the user things based on who they are
  - Campaign – A customization with a targeted focus, usually time limited
  - These are progressively deeper levels of personalization

# MetaTagger and Personalization

- ◆ Personalization frequently depends heavily on categorization of content
  - Geography
  - Language
  - Topic, etc



## Using MetaTagger to Categorize Content

- ◆ MetaTagger is an excellent tool for automated categorization
- ◆ After "training" MetaTagger with a customized vocabulary, it can scan each new or changed file in TeamSite and tag the file with metadata that contains category information
- ◆ This data can then be deployed to your search database by DataDeploy

## Creating "My Portal" Capabilities

- ◆ If you're running a portal server based on WebLogic or another J2EE portal product, TeamPortal can be used to integrate TeamSite features into your portal site
- ◆ It can add a "My TeamSite" and "My Tasks" portlet, which allows access to your TeamSite workarea, tasks, and templates directly via your portal site, bypassing the normal TeamSite UI

# Action Exercise: Personalization Requirements

- ◆ Within your group, spend 20 minutes discussing the following:
  - What personalization requirements does your agency site require?
- ◆ After discussion, spend 20 minutes documenting the following:
  - Your personalization requirements, and how they impact your metadata requirements
- ◆ Each group will then present their findings to the class one at a time

## Action Item Discussion

- ◆ Class presentation
- ◆ Questions
- ◆ Take a few minutes to consider other group action items
  - Integrate theirs with yours if needed



## **Seminar Conclusion**